

Title (en)

METHOD AND APPARATUS FOR GENERATING FIXED-POINT QUANTIZED NEURAL NETWORK

Title (de)

VERFAHREN UND VORRICHTUNG ZUR ERZEUGUNG VON FESTKOMMAQUANTISIERTEM NEURONALEM NETZWERK

Title (fr)

PROCÉDÉ ET APPAREIL DE GÉNÉRATION DE RÉSEAU NEURONAL QUANTIFIÉ À POINT FIXE

Publication

EP 3438890 B1 20240110 (EN)

Application

EP 18187235 A 20180803

Priority

KR 20170098999 A 20170804

Abstract (en)

[origin: EP3438890A1] A method of generating a fixed-point quantized neural network includes analyzing a statistical distribution for each channel of floating-point parameter values of feature maps and a kernel for each channel from data of a pre-trained floating-point neural network, determining a fixed-point expression of each of the parameters for each channel statistically covering a distribution range of the floating-point parameter values based on the statistical distribution for each channel, determining fractional lengths of a bias and a weight for each channel among the parameters of the fixed-point expression for each channel based on a result of performing a convolution operation, and generating a fixed-point quantized neural network in which the bias and the weight for each channel have the determined fractional lengths.

IPC 8 full level

G06N 3/084 (2023.01)

CPC (source: CN EP KR US)

G06F 7/483 (2013.01 - US); **G06F 7/49947** (2013.01 - US); **G06F 7/5443** (2013.01 - US); **G06F 7/57** (2013.01 - CN); **G06F 17/153** (2013.01 - KR); **G06N 3/0464** (2023.01 - KR); **G06N 3/048** (2023.01 - KR); **G06N 3/0495** (2023.01 - KR); **G06N 3/063** (2013.01 - EP US); **G06N 3/08** (2013.01 - CN); **G06N 3/096** (2023.01 - KR); **H03M 7/24** (2013.01 - US); **G06F 2207/4824** (2013.01 - US); **G06N 3/084** (2013.01 - EP US)

Cited by

CN113228056A; EP3805994A1; CN109978135A; WO2021068249A1; WO2021179281A1; WO2020160787A1

Designated contracting state (EPC)

AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

DOCDB simple family (publication)

EP 3438890 A1 20190206; EP 3438890 B1 20240110; CN 109389219 A 20190226; CN 109389219 B 20240409; JP 2019032833 A 20190228; JP 7072464 B2 20220520; KR 102601604 B1 20231113; KR 20190014900 A 20190213; US 11588496 B2 20230221; US 2019042948 A1 20190207; US 2023117033 A1 20230420

DOCDB simple family (application)

EP 18187235 A 20180803; CN 201810868446 A 20180801; JP 2018144829 A 20180801; KR 20170098999 A 20170804; US 201816051788 A 20180801; US 202218084948 A 20221220